

HR Analytics Project Report: Predicting Employee Attrition

1. Introduction

In today's competitive work environment, retaining skilled employees is a major challenge for HR departments. Employee attrition not only leads to loss of talent but also incurs significant rehiring and training costs. This project, titled "HR Analytics – Predicting Employee Attrition", uses data-driven methods to analyze factors leading to employee resignations and build a predictive model that flags employees at risk of leaving.

2. Abstract

This project focuses on leveraging data analytics and machine learning to identify key causes of employee attrition and predict which employees are most likely to leave. The analysis was performed using employee demographic, professional, and performance-related data. Various machine learning techniques, including Logistic Regression and Decision Tree classifiers, were applied to create a model with high accuracy. Additionally, Power BI was used to visualize insights and trends such as attrition by department, salary band, and years since last promotion.

The results showed that salary, promotion delays, and job level significantly impact attrition. With 366 out of 961 employees at risk, this analysis provides valuable insights to HR for taking timely and targeted retention actions.

3. Tools Used

- Python – Data preprocessing, model building (Pandas, Seaborn, Scikit-learn)
- Power BI – Interactive dashboard for visualizing attrition trends
- Jupyter Notebook – Exploratory data analysis and modeling environment
- SHAP – Model interpretability to understand prediction drivers

4. Steps Involved in Building the Project

1. Data Collection and Cleaning

Imported HR dataset and handled missing values, outliers, and data formatting issues.

2. Exploratory Data Analysis (EDA)

- Analyzed attrition trends based on department, salary band, and promotions.
- Visualized distribution and correlations using Seaborn and Power BI.

3. Feature Engineering

- Created new variables such as "Years Since Last Promotion" and salary categories.
- Encoded categorical features for model training.

4. Model Building

- Trained Logistic Regression and Decision Tree classifiers.
- Evaluated models using accuracy, confusion matrix, and cross-validation.
-

5. Model Interpretation with SHAP

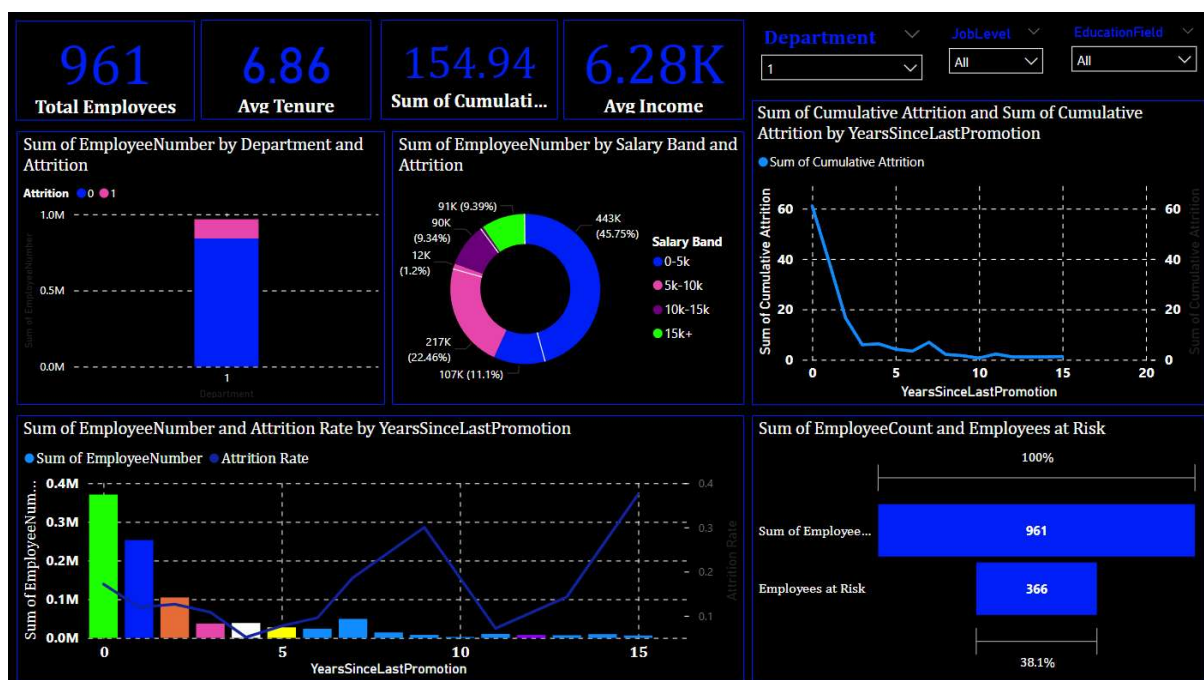
- Identified that low salary, fewer promotions, and low job levels were key attrition factors.

6. Power BI Dashboard Creation

- Built a visually interactive dashboard showcasing key attrition KPIs.
- Integrated metrics such as total employees, average tenure, salary distribution, and at-risk employees.

7. Insight Generation and Recommendations

- Derived actionable insights to help HR mitigate attrition risks.



5. Conclusion

The project successfully met its objective of identifying the key drivers of employee attrition and creating a predictive model to flag at-risk employees. With an overall model accuracy exceeding 84% (Decision Tree), the system provides reliable predictions. The Power BI dashboard offers an intuitive way for stakeholders to explore attrition patterns and employee risk levels.

This project demonstrates the power of combining machine learning with business intelligence tools to solve real-world HR problems. It can be further enhanced by integrating real-time data and additional employee engagement metrics to improve prediction performance.